Translation Studies, Autumn 2024, Vol. 22, No. 87, pp. 110–132: Scientific Research Paper
DOR: 20.1001.1.17350212.1403.22.3.8.5

A Comparative Study on Translation of Persian Colloquialism into English by ChatGPT and Other Translation Platforms[1]

Zeynab Mirhashemi[2], Mahvash Gholami[3] & Hossein Bahri[4]

Abstract

Translating colloquialism as a basic challenge in the process of literary translation could be used as a criteria to investigate the capabilities of the translation machines and tools, most of which use artificial intelligence (AI). The present research was an attempt to investigate the performance of Yandex Translate (YT), Microsoft Bing Translate (BT), Google Translate (GT), ChatGPT, and MateCat (MC), in translating Persian colloquialism. In this process, the researchers tried to compare these platforms' translations, demonstrate their weaknesses and propose improving suggestions to the designers of translation platforms. To this end 202 Persian sentences or phrases containing 240 colloquial expressions, words, and tones were entered into these five platforms and their translations were evaluated based on parameters including semantic accuracy and colloquial language recognition and stylistic transference. Orlando's (2011) grid descriptor was adopted to give grades to the translations and Fuzzy-Math method was used for the precise analysis and comparison of results. In the end, the results revealed the higher position of Microsoft Bing Translate in translating Persian colloquialism.

Keywords. Artificial intelligence, Colloquial language, Fuzzy-math, Machine translation, Stylistic equivalence

---

1. This paper was received on 19.07.2024 and approved on 14.10.2024.

2. M.A. in English Translation Studies, Department of Foreign Language, Faculty of Literature and Humanities, Kharazmi University, Tehran, Iran; email: zeynabmirhashemi7@gmail.com

3. Corresponding Author: Assistant Professor, Department of Foreign Language, Faculty of Literature and Humanities, Kharazmi University, Tehran, Iran; email: mahvashgholami@khu.ac.ir

4. Assistant Professor, Department of Foreign Language, Faculty of Literature and Humanities, Kharazmi University, Tehran, Iran; email: hbahri@khu.ac.ir

1. Introduction

There is no doubt that translation of literary texts plays a prominent role in strengthening cultural awareness in nations which use different languages, in other words, "a translated literary work informs the reader about the foreign literature and culture and it also develops, enhances and enriches the reader's culture" (Karjagdiu & Mrasori, 2021). However, translating literary texts involves complexities which make the task difficult. One of these problems is related to translating colloquial language or colloquialism which is a challenging and effortful task for both human and machines. The present research aimed to assess the quality of translation of colloquialism by five different translation platforms including Yandex, MateCat, Google Translate, Microsoft Bing Translate, and ChatGPT which are mostly artificial-intelligence-driven. Among these, MateCat is somehow different and is in fact a CAT tool which is used in translating texts. ChatGTP is a new Chabot that has human-like conversation abilities and appeared in 2022. It is capable of answering almost all the questions and is helpful in performing different tasks including translation.

The focus of this research was first on whether ChatGPT and other platforms can process and understand the meaning of Persian colloquial words and expressions—semantic accuracy—and the second concern was their ability in recognizing colloquial language and transferring it into target language—stylistic equivalence.

The assessment included explanations of the accomplished results based on House's (1997) functional equivalence model and Popovic's (1976 as cited in Muzaffar & Behera, 2017) theory on stylistic aspect of literary works. The evaluation of data was done through a model by Orlando (2011) whose grid descriptor—as presented in Appendix A—was used to give grades to the translations. In the next stage, Fuzzy-Math (FM) model, proposed by Lotfi Asker Zadeh (1965 as mentioned in Liu & Zhao, 2015) was used for comparison of results.

The present investigation is significant in three aspects: dealing with translation abilities of ChatGPT, focusing on the relatively less-studied subject of translation of colloquial language, and using multiple instruments for analysis of results, i.e., the Functional Equivalence of House (1997), Stylistic theory of Popovic (1976 cited in Muzaffar & Behera, 2017), and the Fuzzy-Math of Lotfi Asker Zadeh (1965 as mentioned in Liu & Zhao, 2015).

The purpose of the research was comparing the performance of the above-mentioned platforms in translating Persian colloquialism. To this end the study addresses the following questions:

Q_1. Which platform among Yandex, MateCat, Microsoft Bing, Google Translate, and ChatGPT has better performance in transferring the semantic and stylistic features of Persian colloquialism when translating literary texts into English?

Q_2. What are these platforms' weaknesses in recognizing and transferring colloquial language?

Q_3. Is there any significant difference among the platforms' quality of translation of Persian colloquialism?

Regarding the third questions the following null hypothesis was formed:

$H_{01.}$ There are no significant difference among the mentioned platforms' quality of translation of Persian colloquialism.

2. Review of the Related Literature

Baldick (2008) defines colloquialism or colloquial language as "the use of informal expressions appropriate to everyday speech rather than to the formality of writing, and differing in pronunciation, vocabulary, or grammar" (p. 61). Translating colloquial expressions are challenging for translators because of their usually weak structure and culture-bound meanings. By appearance of computers and various AI tools for translation of foreign languages, everyone expected their good performance in translating any type of text, but academic researches on translation quality assessment have questioned their abilities (e.g. Aghai, 2024;

Almahasees, 2018; Chochiang et al., 2020; Sutrisno, 2020). However, the translation quality of these tools and machines varies and some of them are considered acceptable in comparison with others.

One of the widely used machine translations is Google Translate (GT) which started in April 2006. It performed poorly till 2016 when GT's model changed from SMT (Statistic Machine Translation) to NMT (Neural Machine Translation). The recent statistics show that GT performs more accurately in translating from and into major languages including English, French, German, Spanish, and Chinese (Moltzau, 2020). Today it supports more than one hundred languages and is being used by hundreds of millions of users (Turovsky, 2016).

Microsoft Bing Translate (BT) is the earliest MT whose development goes back to 2003 when the only languages it could support were English into Spanish, French, German, and Japanese. After the appearance of AI technology, BT started to use NMT instead of SMT resulting in more accurate and fluent translations. Today, more than one hundred languages are covered by BT and billions of users use its services (Mohan & Skotdal, 2021).

There have been higher numbers of research on GT and BT. Almahasees (2018) performed a comparative analysis between GT and BT in translating journalistic texts in Arabic. Considering grammatical and lexical accuracy, GT with the rate of 79.8% and BT with the rate of 74.5% achieved over 90 percentage of accuracy. Another research was done by Sutrisno (2020) on GT in translating English to Indonesian and the result was 60.37% accuracy in translating words and phrases which was not what they expected and "[…] much higher than that of other Asian languages as reported by Patil and Davis (2014), which averaged 46% accuracy".

Yandex, a lesser-known MT, was released in 2011 for English, Russian, and Ukrainian. Today, it translates from and into more than a hundred languages and has almost thirty millions users (Yandex AI, n.d.). YT uses both systems of NMT and

SMT and is considered a hybrid machine.

Chochiang, Thongkhamdee, and Sathansat (2020) implemented a comparative research on GT, Baidu Translate, Yandex Translate and BT, comparing their translations of social media comments from English to Thai. The results showed that GT is the most suitable platform with the rate of 89.33% but it was limited in producing appropriate words. Yandex and BT were found to be incompetent (Chochiang et al. 2020).

Another investigation was performed on translations from Bahasa into English, examining quality of translation of news items by Yandex for which the results reflected its competence (Sumasjo & Mahanani, 2020).

MateCat (MC) is an online Computer Assisted Translation (CAT) tool which was presented in 2015. Although some people consider it as a machine translation and use it for translating texts, CAT acts like a database and is not programmed to translate texts by itself. MateCat covers ninety languages and uses the largest translation memory, which contains twelve billion words (MateCat, n.d.). Its specific features, such as different file formats, project settings, statistics, editor, and answering support, make its application easy (Akhrameev, 2015).

Bououden and Saida (2022) have performed a comparative investigation between translation of a scientific text from English to Arabic by GT and MC and the results showed the outperformance of MC over GT. Another research was conducted by Cornet and De Keizer. (2017) comparing GT, MC, and Thot in translating Dutch terminologies. It concluded that GT and MC translated 85.4% similarly and both outperformed Thot.

ChatGPT differs from the above mentioned platforms in one aspect. As the most recent technology offered by OpenAI, it is not merely an Artificial-Intelligence-based system, rather a communicator which answers questions and acts as a consultant in conversations. It could also be used as a machine translation which translates and communicates in a conversational way – it is chat-based (OpenAI, 2022). ChatGPT applies a massive language data and is able to translate among over ninety languages. Its superiority over other MTs is that it translates more naturally, and translators can use it to create connections not only among languages but also among cultures (Frąckiewicz, 2023). As it is chat-based, ChatGPT can recognize and correct errors or provide offers for better alternatives. Another prominent feature is its speed in producing outputs. The system utilizes Natural Language Processing (NLP) technology which amalgamates the methods of Rule Based Machine Translation (RBMT), SMT, and Deep Learning (DL) to understand the meaning of the text and intention of the author (IBM, n.d.); therefore, its translations can seem to be more like human-produced ones (Choudhury, 2023; Frąckiewicz, 2023).

The research on ChatGPT includes a comparative study by Aghai (2024). His study has assessed the quality of literary translation from Persian to English using ChatGPT and Google Translate and concluded that both are similar and have limitations in respect to accuracy, equivalence, and text function. Another research is one by Khoshafah (2023) who has examined ChatGPT in translating Arabic texts into English and the results suggested it is suitable for simple content but not for complex texts, such as legal documents, medical reports, scientific studies, and literary works.

Stap and Araabi (2023) have conducted a contrastive analysis among different systems using NLP, examining translations of a Spanish text into 11 indigenous languages in South America – which are low-resource languages ; low-resource languages are those that have relatively less data available for training conversational AI systems (Magueresse, Carles & Heetderk, 2023). The results showed that although widely used, ChatGPT does not perform efficiently in the case of low-resource languages. Shamsfard (2019) claims that Persian, too, is considered a low-resource language. Another research was done by Hendy et al. (2023) on different low and high resource languages applying different models of ChatGPT including ChatGPT, GPT3.5 (text-davinci-003), and text-davinci-002. They also concluded that ChatGPT is limited in translating low-resource languages compared to high resource ones.

3. Methodology

The purpose of the research was comparing the performances of the above-mentioned platforms in translating Persian colloquialism. To this end this study followed a descriptive-analytic design and in the process of evaluation the following models were used:

1. The models for semantic and stylistic equivalence presented respectively by House (1997) and Popovic (1976 as cited in Muzaffar & Behera, 2017) were utilized as parameters against which the target sentences were examined and the elements each contained were used as follows: in terms of semantic equivalence, the main meaning of the source sentences was concerned, and in terms of stylistic equivalence, colloquial tone, colloquial expressions, and colloquial words were focused on. The criterion was how the platforms comply with these parameters; the compliance was divided into two categories: colloquialism recognition and transference. In terms of semantic accuracy both recognition and transference could naturally occur at the same time; in terms of stylistic equivalence the platforms could either recognize the colloquialisms and transfer the source items or only recognize them – as will be explained later.

2. The method of Fuzzy-Math proposed by Lotfi Asker Zedeh (1965 as mentioned in Liu & Zhao, 2015) was used to attribute weight to each parameter and evaluate the qualities based on the results achieved by counting A to E marks of each expression. The processes needed to calculate the final degree of the translations' qualities according to each parameter. The evaluative framework adopted the process followed by Liu and Zhao (2015) combining House's model and Fuzzy-Math.

3.1. The Corpus of the Study

The corpus of the research included 240 colloquial expressions, words and tones selected from Persian translation of the novel *The Sound and the Fury* translated by Bahman Sholehvar (1974). The reason for choosing a translated text rather than an original Persian text as corpus was that the selected text is a rich source of Persian colloquial elements and as some theorists (e.g Bassnett, 2003, p. 72; Bassnet, 2006; Bush & Bassnet, 2006) believe, translator is a writer. "The translator is both a reader and a writer and, therefore, there is no such thing as the dichotomy translator/translation (target text) and author/original (source text)" (Silveira-Brisolara, 2011 p. 120).

4. Data Collection and Analysis

Three pages in each chapter were chosen randomly through a random number generator website—*random.org*—and the repeated numbers were ignored; four sentences from each page were selected purposefully, and the final corpus contained 202 sentences or phrases which included 240 colloquial items.

In identifying the colloquial terms McCrimmon's (2020, p. 167) model was used. Among the items he mentions as features of colloquialism the following could be identified in Persian colloquialism: relatively short simple sentences, often grammatically incomplete, with few rhetorical devices; use of contractions, clipped words, and the omission of pronouns which would be retained in a formal style; a vocabulary marked by general avoidance of learned words and by inclusion of some less objectionable slang terms; a simplified grammatical structure which leans heavily on idiomatic constructions and sometimes ignores the fine distinctions of formal grammar and; a personal or familiar tone, which tries to create the impression of speaking intimately to the reader.

Then the selected expressions were translated by five platforms: GT, BT, YT, MC, and GPT. The reason behind choosing these five translation platforms out of other free ones was that, based on a brief pilot study, their performance in translating Persian colloquial language was more acceptable. After the data was gathered one of the researchers evaluated the output—containing two hundred forty colloquial items—and another evaluator, a BA graduate of English Translation

Studies, evaluated a random quarter of the target text (TT)—containing sixty colloquial items—making use of rating scale of Fuzzy-Math. To have an objective view, the Pearson correlation was calculated to see through the reliability of the general evaluation and the grid descriptor. The correlation of their evaluations was 0.8234 indicating a strong positive correlation at 0.01.

The prominent goal of the research was evaluating and assessing each platform's performance separately while considering their adherence to semantic and stylistic features of the source text, and comparing their performances. To these ends, the researchers used Fuzzy-Math to analyze the outputs and semantic and stylistic equivalences to elaborate on the evaluation results.

In the process of evaluation, the researchers adopted a grid descriptor designed by Orlando (2011). It was changed in some parts to fit the goals and the corpus of the present research (refer to Appendix A). The grid descriptor contains three parts—three general dimensions—including 1) semantic recognition and transference, 2) colloquial recognition and 3) stylistic transference. Each dimension includes criterion for every grade; A (very good) … and E (very bad). The selected items for evaluation—related to sub-dimensions of tone, expression, and word—were those which were not repetitive, and also their removing or changing would influence the meaning.

To clarify the difference between recognition and transference of colloquial words or expressions, consider the expression زور بده. All of the five platforms recognized the colloquial expression of 'زور دادن' or 'زور آوردن' which consists of two words of زور (force or power) and دادن (give); however, only Microsoft, Google, and MateCat could transfer the colloquial style of it with the word 'push'. Yandex and ChatGPT used the more formal word 'forcing' (with the grade of C for stylistic transference) or 'applying force' (with the grade of D for stylistic transference) which is even more formal.

The last point in the evaluation was the level at which the evaluation of the source text items occurred, i.e. three levels of sentence, phrase, and word.

The next step was the application of the Fuzzy formula including dimensions' weights and grades, and the matrix presented by Liu and Zhao (2015). After the operation, the achieved numbers were collected, and the results were shown from highest to lowest number e.g. 32.6(A)>17(E)>7.8(D)>5.2(C)>2.4(B) which were indicators of the quality of each platform's performance.

5. Results

The grades of each platform in every general dimension (semantic recognition and transference, colloquial recognition and stylistic transference) and the sub-dimensions (tone, expression and word) were counted. Table 1 shows examples for grading the translations of each platform.

Table 1. *Examples for Grading the Translations*

Colloquial expression	Yandex	Microsoft	Google	ChatGPT	Matecat
تا بتوانند آنرا به یک مشت یانکی بدهند	So they can give it a Yankee punch	so that they could give it to a handful of Yankees	so they can give it with a Yankee fist	Until they can give it a good Yankee punch	So they can give it back to them
Semantic grade	E	A	E	E	C
Stylistic grade	E	A	D	D	D
من دلم واسه بابا جون ملوسم تنگ شده"	I miss my father	I miss my sweet daddy	I miss my father Jun Melosem	I miss dear dad	I miss Santa Claus
Semantic grade	B	A	D	B	E
Stylistic grade	B	A	D	B	E
مطلب دستگیرم شد	I was arrested	I got it	I was arrested	The issue caught my attention	Item seized
Semantic grade	E	A	E	B	E
Stylistic grade	E	A	E	C	E

The total results are reflected in Table 2. In this table, for example, Yandex in semantic recognition and transference has got grade A in 55 cases, grade B in 14 cases, grade C in 27 cases, grade D in 24 cases and grade E in 82 cases.

Table 2. *Grades of the five platforms for each dimension and sub-dimension*

Platforms / Dimensions	Yandex					Microsoft					Google Translate					ChatGPT					MateCat				
Semantic (0.5) 202 samples 240 elements	55A	14B	27C	24D	82E	120A	15B	19C	14D	34E	77A	14B	17C	19D	75E	67A	17B	18C	21D	79E	41A	6B	20C	27D	108E
Total colloquial recognition (0.3)	101A	4B	8C	8D	119E	164A	4B	10C	2D	60E	109A	6B	9C	9D	107E	101A	4B	11C	16D	108E	73A	2B	5C	10D	150E
Colloquial tone	45A	0B	2C	1D	24E	53A	1B	1C	0D	17E	43A	1B	1C	4D	23E	43A	1B	1C	4D	23E	27A	1B	1C	4D	39E
Colloquial expression	28A	2B	2C	7D	51E	57A	3B	6C	1D	23E	34A	3B	4C	5D	44E	28A	2B	7C	8D	45E	25A	0B	2C	4D	59E
Colloquial word	28A	2B	4C	0D	44E	54A	0B	3C	1D	20E	32A	2B	4C	0D	40E	30A	1B	3C	4D	40E	21A	1B	2C	2D	52E
Total Stylistic transference (0.2)	91A	11B	7C	12D	19E	160A	13B	4C	3D	60E	96A	13B	15C	7D	109E	74A	19B	24C	17D	106E	63A	5B	12C	9D	151E
Colloquial tone	42A	2B	2C	2D	24E	53A	2B	0C	0D	17E	38A	3B	4C	4D	23E	36A	5B	5C	3D	23E	25A	1B	4C	3D	39E
Colloquial expression	26A	3B	1C	8D	52E	56A	6B	3C	2D	23E	30A	5B	8C	3D	44E	17A	7B	12C	10D	44E	19A	1B	5C	5D	60E
Colloquial word	23A	6B	4C	2D	43E	51A	5B	1C	1D	20E	28A	5B	3C	0D	42E	21A	7B	7C	4D	39E	19A	3B	3C	1D	52E

The contents of this table were summarized in Table 3 using Lotfi Asker Zadeh's Fuzzy-Math model as presented in Liu and Zhao (2015). The formula is a simple operation of counting each platforms' grades—A separately, B separately, etc.—in each dimension and then multiplying the results by their determined weights. The weights proposed by Liu and Zhao (2015) were used due to the suitability to the present research's goal of importance hierarchy – 0.5 for semantic, 0.3 for colloquial recognition and its sub-dimensions, and 0.2 for stylistic transference and its sub-dimensions. For example in the case of Yandex it had achieved 55 'A's for semantic equivalence, 101 'A's for colloquial recognition and 91 'A's for stylistic transference. When 55 was multiplied by 0.5 (= 27.5) plus 101 multiplied by 0.3 (= 30.3) plus 91 multiplied by 0.2 (=18.2) the result was 76. They were added to each other to reach to the general dimensions' of each platforms rank as is reflected in Table 3.

Table 3. *Performance of Each Platform on General Dimensions*

platforms / Grades	Yandex	Microsoft	Google	ChatGPT	MateCat
A	76	141.2	90.4	78.6	55
B	10.4	11.3	11.4	13.5	4.6
C	17.3	13.3	14.2	17.1	13.9
D	16.8	8.2	14	18.7	18.3
E	100.5	47	91.4	93.1	129.2
Total	221	221	221	221	221
Comparison	100.5(E)>76(A)	141.2(A)>47(E)	91.4(E)>90.4(A)	93.1(E)>78.6(A)	129.2(E)>55(A)
Result	E	A	E	E	E

Different levels of adherence to the semantic equivalence of House (1997) and the stylistic equivalence of Popovic (1976 as cited in Muzaffar & Behera, 2017) were the criteria based on which the analyses were performed.

The first question was: Which platform among Yandex, MateCat, Microsoft Bing, Google Translate, and ChatGPT has better performance in transferring the semantic and stylistic features of Persian colloquialism when translating literary texts into English?

The analysis of data indicated that considering general dimensions, Microsoft performed better than the other four. This is reflected in Table 3 which shows Microsoft superiority over the other four in performing general dimensions' translations. It achieved 141.2 out of 221 points for 'A' which is an indicator that it received the highest grade among others—a generally very good performance in translating Persian colloquialism. The lowest grades belong to MateCat and Yandex with 55 and 76 out of 221 points respectively for grade A.

The second question was: What are these platforms' weaknesses in recognizing and transferring colloquial language?

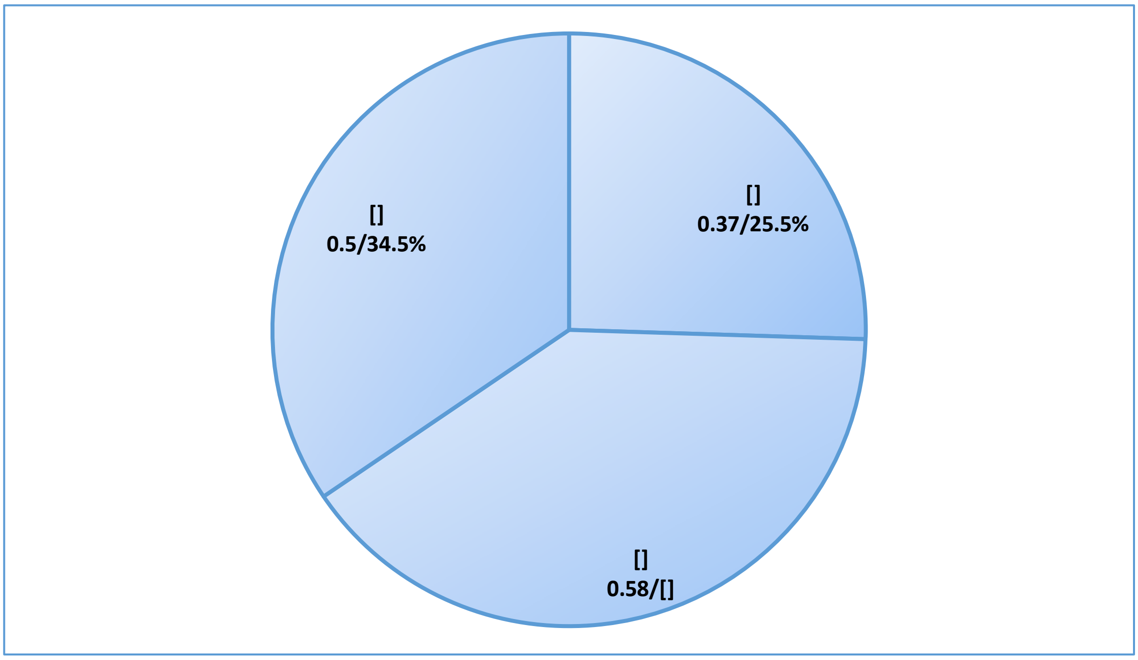
The following two samples are the representatives of good recognition of colloquialism or both recognition and transference of stylistic elements and very bad performance in semantic dimension.

The translation of the sample "توام تی پی سربسر کوئینی نذر، میشنفی" is "You can hear the vow of Queenie's head", and the other sample is " کدی یادت میاد واسه اینکه تنکه‌ات گلی شده بود دیسلی بهت غرغر میکرد" for which the translation is "You remember the code *because* your tank was a flower, and the dyslexia was *grumbling* at you". As it is apparent in Appendix B the platforms were able to recognize and transfer the items such as غرغر میکرد or میشنفی but they could not understand and transfer the general meaning of the sentence.

Another inference is that they are more successful in recognizing semantic elements than in transferring the style, which means they performed better in recognizing the meaning of the elements than transferring them in an equivalent informality. The following sample reflects this inference: "باز دیسلی داره *بددهنی* میکنه" is translated "He/she is being *rude and disrespectful*". As the example shows, the meaning is recognized, but the informal style of the item is not preserved.

Chart 1 reflects the summary of the analysis for answering the second research question. The first numbers are Fuzzy Math calculations of 'A' grades of general dimensions. The second numbers are percentage of the first numbers out of 100%.

Chart 1. *Platforms' Performance in General Dimensions by Proportion of 'A's*



Note. Proportion of 'A's = collection of 'A's of the five platforms' general dimensions divided by the total numbers of the samples multiplied by five

The following table reflects a summary of analyses of the platforms' performance in colloquial recognizing and transferring sub-dimensions.

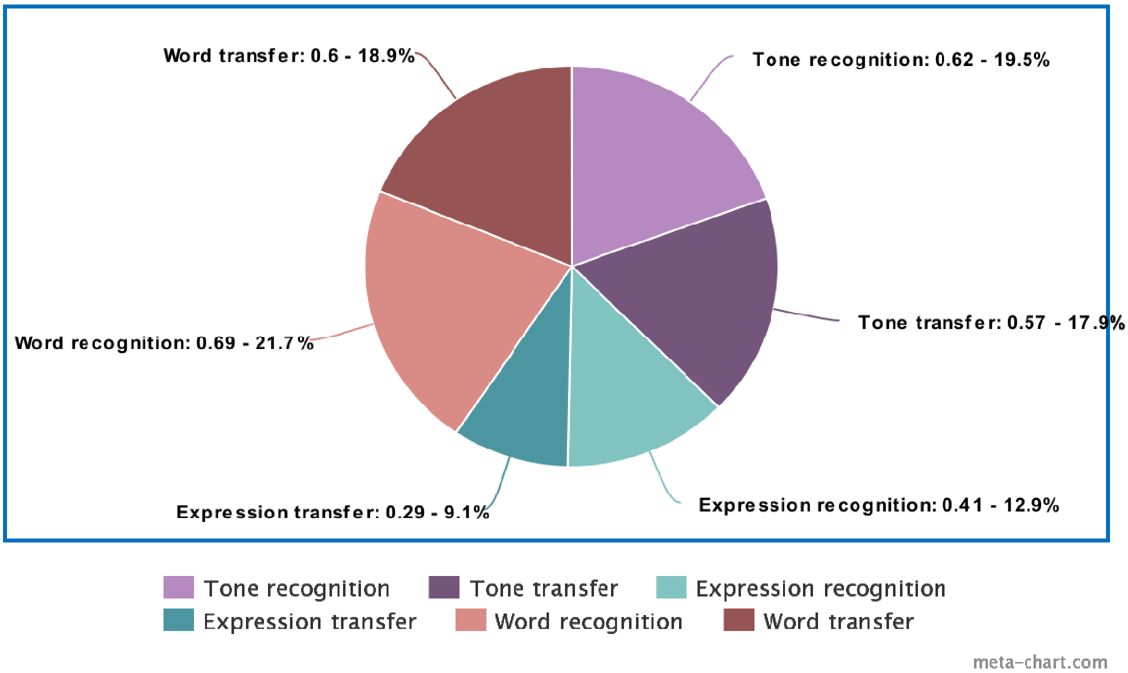
Table 4. *The Results of the Analysis of Platforms' Performances in Sub-Dimensions*

	Yandex			Microsoft			Google			ChatGPT			Matecat		
Grades	expression	word	tone	expression	word	tone	expression	word	tone	expression	word	tone	expression	word	tone
A	13.6	13	21.9	28.3	26.4	26.5	16.2	15.2	20.5	11.8	13.2	20.1	11.3	10.1	13.1
B	1.2	1.8	0.4	2.1	1	0.7	1.9	1.6	0.9	2	1.7	1.3	0.2	0.6	0.5
C	0.8	2	1	2.4	1.1	0.3	2.8	1.8	1.1	4.5	2.3	1.3	1.6	1.2	1.1
D	3.7	0.4	0.7	0.7	0.5	0	2.1	0	2	4.4	2	1.8	2.2	0.8	1.8
E	25.7	21.8	12	11.5	10	8.5	22	20.4	11.5	22.3	19.8	11.5	29.7	26	19.5
Total	45	39	36	45	39	36	45	39	36	45	39	36	45	39	36
Comparison	25.7 > 13.6	21.8> 13	21.9> 12	28.3 > 11.5	26.4> 10	26.5 > 8.5	22>16.2	20.4>15.2	20.5)>11.5	22.3>11.8	19.8>13.2	20.1>11.5	29.7>11.3	26>10.1	9.5>13.1
Result	E	E	A	A	A	A	E	E	A	E	E	A	E	E	E

It is evident that only Microsoft Bing has achieved 'A' grade in all three sub-dimensions and the worst performance belongs to MateCat.

The following chart shows the comparison of platforms' performance in sub-dimensions when the proportion of 'A' grades are the criteria. The calculations has been done using Fuzzy-Math

Chart 2. *Platforms' Performance in Sub-Dimensions by Proportionality of 'A's*



The general analysis of sub-dimensions of stylistic dimension, as reflected in chart 2, indicates that the highest grade of 0.69 was achieved for word recognition and the lowest grade of 0.29 was achieved for expression transfer which indicated a very good performance towards colloquial word (with the transfer grade of 0.60) and a very bad performance towards colloquial expression (with the recognition grade of 0.41). Also, colloquial tone's recognition and transfer were at average level with the grades 0.57 and 0.62. Appendix B includes translation samples which show platforms' performance in sub-dimensions.

Therefore, within the limited scope of the present research, the best general performance was acted toward stylistic recognition and the lowest was regarding semantic; the weak performance regarding the total sentence meaning or semantic is an indicator of weakness in context comprehension – in many cases when there were wrong translations, the contexts were added, and the results were still the same. Finally, tone recognition and tone transference gained the highest percentage which demonstrate the platforms' good performance regarding translating Persian colloquial tone. The lowest percentages belong to word transference and expression transference; the platforms recognize Persian colloquial language better than they transfer its forms.

The third question was: Is there any significant difference among the platforms' quality of translation of Persian colloquialism?

To answer the last question and reject or support its null hypothesis, the researchers performed an ANOVA by SPSS 26 and the result is shown in Table 5.

Table 5. *Total ANOVA Calculation*

	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	3308.700	4	827.175	8.439	.001
Within Groups	1470.250	15	98.017		
Total	4778.950	19			

As it is reflected in Table 5, the significance level is 0.001, which is less than 0.05, indicating that the differences between the means are statistically significant. Therefore, the answer to the third question is yes there is a significant difference among the platforms' quality of translation of Persian colloquialism. This means a rejection of the null hypothesis.

6. Discussions

The substitution of humans with computers has triggered research in all fields including the translation industry. Results of the researches on machine translations (e.g. Chochiang et al., 2020; Bououden and Saida, 2022; Almahasees, 2018; Cornet et al., 2017 and Sutrisn, 2020) has shown that these tools are mostly good translators of special texts (e.g. news items, scientific texts, journalistic texts and terminologies); however, they fail in translating non-standard colloquial language such as idiomatic expressions (Musaad & Al Towity, 2023). The failure of these tools could be related to two aspects.

The first reason could be the great differences between the languages involved. It is evident that the mistakes are less when the two languages are somehow similar, such as English, French, German or Spanish. Another reason could be related to the insufficiency of the corpus which forms the basis for programming the platforms.

The appearance of AI and later, ChatGPT has presented new chances for MTs and tools. Timothy (2023) has discussed the comparison between GT and ChatGPT in terms of language coverage. Google Translate supports more than 140 languages, while ChatGPT is believed to cover even more. He found out that GT performed more successfully than ChatGPT. Also, ChatGPT has not performed efficiently in the case of low-resource languages (Cornet, et al, 2017). The findings of the present study too confirm this fact – if Persian is considered a low-resource language, as it is claimed by Shamsfard (2019).

On the other hand, the results of the present research is against those which mostly mentioned GT as the most appropriate MT. It is to be noted that some of the researches did not consider the language form. In the present study, generally, the platforms showed a more adequate performance towards semantic dimension and lesser one towards stylistic transference. Categorically, word recognition with the percentage of 21.7 gained the highest grade. In contrast, expression transfer with the percentage of 9.1 is at the lowest level.

Additionally, the problem is not just transferring meaning but style and tone are also involved; however, with the context provided, ChatGPT can adjust elements like register, terminology, and style to cater to the audience and purpose (Siu, 2023), but the present study showed that ChatGPT lay behind BT in translating Persian colloquialism. The reason could be the fact that they lack human cultural and emotional senses and mastery of non-standard language and colloquialism of Persian. Hence, human translators are essential to supervise platforms' outputs to ensure culturally appropriate, high-quality translations (Siu, 2023).

Lack of related corpora could also be a reason for the failure. So far, Persian colloquialism does not possess a sufficient corpus which can be the reason why MT and AI face difficulties in Persian colloquialism translation (Khojasteh et al., 2020).

There have been several attempts to improve Persian colloquial corpora (e.g. Khojasteh et al., 2020; Rabiei et al., 2023). Removing these obstacles could help AI and the related tools to gain more power in translating colloquialism.

7. Conclusion

Eventually, based on the results achieved from data analysis, it could be concluded that platforms in translating Persian colloquialism are generally acceptable. To be more specific, BT gained more 'A's than the others which make it the most proper platform regarding informal language among the five. In view of the results, corpus designers should pay more attention to Persian colloquialism to provide a more comprehensive source for MT and AI designers. On the other hand, the designers also should reconsider the algorithms in neural systems to reach a more thorough linkage to low-resource languages. At the end, since the present study is limited to a few aspects of informal language and also to a single language, interested researchers can focus on other aspects of colloquialism in other languages. Finally, they can make use of a multiply of instruments such as Fuzzy-Math and the grid descriptor offered by this research and apply them on other languages to discover other roots of translation platforms' shortages or to assess other platforms' functions.

Works Cited:

Aghai, M. (2024). ChatGPT vs. Google Translate: Comparative analysis of translation quality. *Iranian Journal of Translation Studie*s, 22 (85), 87–103.

Akhrameev, S. (2015). MateCat review: Free online CaT tool for freelance translators. *Medium.com.* Retrieved April 16, 2024 from: https://medium.com/@RussianTranslatorPro/MateCat-review-free-online-cat-tool-for-freelance-translators-44e24663381.

Almahasees, Z. (2018). Assessment of Google and Microsoft Bing translation of journalistic texts. *International Journal of Languages, Literature and Linguistics,* 4 (3), 231–236.

Baldick, C. (2008). *Oxford dictionary of literary terms*. OUP

Bassnett, S. (2003). *Estudos de tradução: Fundamentos de uma disciplina* [Translation Studies: Fundamentals of a discipline. Translated by Vivina de Campos Figueiredo]. Lisbon: Fundação Calouste Gulbenkian

Bassnett, S. (2006). Writing and Translating. In S. Bassnet & P. Bush (Eds.), *Translator as writer* (pp. 173–183). Continuum.

Bassnett, S & Bush, P. (Eds.) (2006). *Translator as writer*. Continuum.

Bououden, R., & Saida, K. (2022). Comparing the effectiveness of Google Translate and MateCat tools in the translation of scientific texts from English into Arabic. *In translation (فی الترجمه)*, 9 (1), 543–560.

Bush, P. & Bassnett, S. (2006). Introduction. In S. Bassnet & P. Bush (Eds.), *Translator as writer* (pp. 1–8).Continuum.

Chochiang, K., Thongkhamdee, T., & Sathansat, L. (2020). Translation quality assessment of online translation systems in translating English to Thai on Phuket Tourism. *Journal of Computer Science*, 16(11), 1535–1545. https://doi.org/10.3844/jcssp.2020.1535.1545

Choudhury, S. (2023). Languages supported by ChatGPT and how to use it in other languages. *MLYearning.org*. Retrieved May 18, 2024 from: https://www.mlyearning.org/languages-supported-by-chatgpt/

Cornet, R., Hill, C., & De Keizer, N. (2017). Comparison of three English-to-Dutch machine translations of SNOMED CT procedures. *PubMed*, *245*, 848–852. https://pubmed.ncbi.nlm.nih.gov/29295219

Frąckiewicz, M. (2023, April 5). The role of ChatGPT-3.5 in advancing machine translation and language localization. *Ts2 Space*. Retrieved April 10, 2024 from: https://ts2.space/en/the-role-of-chatgpt-3-5-in-advancing-machine-translation-and-language-localization/#gsc.tab=0

Hendy, A., Abdelrehim, M., Sharaf, A., Raunak, V., Gabr, M., Matsushita, H., Kim, Y. J., Afify, M. & Avadalla, H. H. (2023). How good are GPT models at machine Translation? A comprehensive evaluation. *arXiv.org*. Retrieved March 19, 2024 from: https://arxiv.org/abs/2302.09210

House, J. (1997). *Translation quality assessment: A model revisited.* Tubingen: G. Narr.

House, J. (2015). *Translation quality assessment: Past and present.* Routledge.

IBM. (n.d.). What is Natural Language Processing? *IBM.* Retrieved March 10, 2024 from: https://www.ibm.com/topics/naural-language-processing

OpenAI. (2022). Introducing ChatGPT. Retrieved March 2, 2024 from: https://openai.com/index/chatgpt/

Karjagdiu, L. & Mrasori, N. (2021). The role of literary translation in the development and enrichment of national Literature. *Journal of Language and Linguistics Studies,* 17(4), 2332–2345.

Khojasteh, H. A., Ansari, E., Bohlouli, M. (2020). LSCP: Enhanced Large Scale Colloquial Persian Language Understanding. *arXiv.org.* Retrieved May, 2024 from: https://arxiv.org/abs/2003.06499

Khoshafah, F. (2023). ChatGPT for Arabic-English Translation: Evaluating the Accuracy. *europepmc.org.* Retrieved March 15, 2024 from: https://europepmc.org/article/ppr/ppr645536

Liu, X., & Zhao, Y. (2015). The study on Quantitative Evaluation in the Translation Quality Management Based on the House's Translation Quality Assessment Model. *Advances in Social Science, Education and Humanities Research: Proceedings of the 2015 international conference on social science and technology education (April 2015).* DOI: 10.2991/icsste-15.2015.250

Magueresse, A., Carles, V., & Heetderk, E. (2023). Low-resource languages: A review of past work and future challenges. *ArXiv.* Retrieved on February 25, 2024 from: https://www.semanticscholar.org/paper/Low-resourceLanguages%3AAReviewofPastWorkandMagueresseCarles/4de3595439ed8e433e1997b49ea9171c01dc846

MateCat. (n.d.). Main benefits for LSPs and freelance translators. *MateCate.com.* Retrieved on May 12, 2024 from: https://site.MateCat.com/benefits#:~:text=MateCat%20is%20an%20online%20CAT%20tool%20that%27s%20free%20and%20easy%20to%20use&text=It%20is%20and%20will%20always,Unlimited%20users%2C%20projects%20and%20storage

McCrimmon, J. M. (2022). *Writing with a purpose* [Short Edition]. Houghton Mifflin Company.

Mohan, K., & Skotdal, J. (2021). Microsoft Translator: Now translating 100 languages and counting! *Microsoft.com.* Retrieved March 25, 2024 from:

https://www.microsoft.com/en-us/research/blog/microsoft-translator-now-translating-100-languages-and-counting/

Moltzau, A. (2020). The History of Google Translate. *Medium*. Retrieved May 23, 2024 from: https://alexmoltzau.medium.com/the-history-of-google-translate-fcbe9de3c10e

Musaad, D. M. M. A. M., & Al Towity, D. A. A. (2023). Translation Evaluation of Three Machine Translation Systems, with Special References to Idiomatic Expressions. مجلة العلوم التربوية و الدراسات الإنسانية *[Humanities and Educational Science Journal]*, 29, 678–700. Retrieved from: https://hesj.org/ojs/index.php/hesj/article/view/700

Muzaffar, S., Behera, P., Jha, G. N., Hellan, L., & Beermann, D. (2017). The Concepts of Equivalence, Gain and Loss (Divergence) in English- URDU Web-Based Machine Translation Platforms. *Proceedings of International Conference of South Asian Languages (ICOSAL-12)* Retrieved February 12, 2024 from: https://scholar.google.co.kr/citations?view_op=view_citation&hl=ko&user=tekYvP0AAAAJ&citation_for_view=tekYvP0AAAAJ:MXK_kJrjxJIC

Orlando, M. (2014). Evaluation of translations in the training of professional translators. *Interpreter and Translator Trainer*, 5 (2) 293–308. Published online: 10 Feb 2014. https://doi.org/10.1080/13556509.2011.10798822

Patil, S & Davis, P. (2014). Use of Google Translate in medical communication: Evaluation of accuracy. *British Medical Journal*. doi: 10.1136/bmj.g7392 (Published 15 December 2014)

Rabiei, L., Rahmani, F., Khansari, M., Rajabi, Z., Salimi, M. (2023). Colloquial Persian POS (CPPOS) corpus: a novel corpus for Colloquial Persian part of speech tagging. *ArXiv.org*. Retrieved May 17, 2024 from: https://www.semanticscholar.org/paper/Colloquial-Persian-POS-(CPPOS)-Corpus%3A-A-Novel-for-Rabiei-Rahmani/8f4f06888d118d7ec10b3a03dd0160831cb89b89

Shamsfard, M. (2019). Challenges and Opportunities in Processing Low Resource Languages: A study on Persian. *Proceedings of the Language Technologies for All (LT4All)*, pages 291–295. Paris, UNESCO Headquarters, 5–6 December, 2019.

Sholevar, B. (1974). *Khashm-o-Hayahoo* [Sound and Fury]. Tehran: Pyrooz Publishing House.

Siu, S. C. (2023). ChatGPT and GPT-4 for Professional Translators: Exploring the Potential of Large Language models in translation. *Social Science Research Network*. Retrieved January 19, 2024 from:

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4448091. https://doi.org/10.2139/ssrn.4448091

Silveira Brisolara, V. (2011). The translator as an Author. *Nonada: Letras em Revista*, 1 (16), 107–125.

Stap, D. & Araabi, A. (2023). ChatGPT is not a good indigenous translator. *Proceedings of the Workshop on Natural Language Processing for Indigenous Languages of the Americas.* (Published in AMERICASNLP 2023). url: https://api.semanticscholar.org/CorpusID:259833859

Sumasjo, H. P., Mahanani, M. P. (2020). Analysis of Yandex Translate translation quality of new items from Indonesian to English. *The Asian Journal of English Language and Pedagogy*, 8 (1), 1–7. Retrieved February 22, 2024 from: https://ejournal.upsi.edu.my/index.php/AJELP/article/view/3096

Sutrisno, A. (2020). The accuracy and shortcomings of Google Translate translating English sentences to Indonesian. *Education Quarterly Reviews,* 3(4), 555–568. https://doi.org/10.31014/aior.1993.03.04.161

Timothy, M. (2023). ChatGPT vs. Google Translate: Which Is Better At Translation? *MUO*. Retrieved February 14, 2024 from: https://www.makeuseof.com/chatgpt-vs-google-translate-which-is-better-at-translation/

Turovsky, B. (2016). Ten years of Google Translate. *Google*. Retrieved May 4, 2024 from: https://blog.google/products/translate/ten-years-of-google-translate/

Yandex AI. (n.d.). Bridging the Language gap with Neural Translation of Videos, Images and Text. Retrieved February 15, 2024 from: https://ai.yandex.com/blog/neural-translation

فاکنر، ویلیام. ۱۳۵۳.خشم و هیاهو. ترجمه بهمن شعله‌ور. انتشارات پیروز.

Appendices link:

https://drive.google.com/file/d/1cIKGwAw4cja9MERQt5U24gOXEsV9UqZq/view?usp=drive_link